UDC 621.391

ANALYSIS OF CYCLE CONSISTENT GENERATIVE ADVERSARIAL NETWORKS FOR UNPAIRED IMAGE TRANSLATION

YU LIU

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

Received March 31, 2026

Abstract. Unpaired image translation has become an important research topic in computer vision due to the difficulty of obtaining paired datasets in real-world scenarios. Cycle consistent generative adversarial networks (CycleGAN) provide an effective solution by learning mappings between two image domains without paired supervision. This paper presents a comprehensive analysis of the CycleGAN model, including its architecture, training strategy, and key components such as cycle consistency and identity mapping mechanisms. Experiments are conducted on benchmark datasets, including horse-to-zebra and seasonal translation tasks, and evaluated using SSIM and PSNR metrics. The results demonstrate that CycleGAN can generate visually realistic images while effectively preserving the structural content and essential features of the input images. Compared with traditional supervised methods, it significantly reduces the reliance on labeled data while still achieving competitive performance in various scenarios. However, limitations such as insufficient semantic consistency, visual artifacts in complex regions, and limited controllability are also discussed in detail, indicating directions for future improvement.

Keywords: CycleGAN, generative adversarial networks, image translation, deep learning, computer vision.

Introduction

Image-to-image translation is a basic task in computer vision. The goal is to transform an image from one domain into another while keeping the main content unchanged. This problem appears in many real situations, such as changing weather conditions, converting day images to night images, or translating medical images between different types.

Traditional methods depend on paired datasets, where each input image has a corresponding output image. However, such datasets are difficult to collect. In many cases, it is impossible to obtain two images of exactly the same scene under different conditions. This limits the use of supervised learning methods.

Generative adversarial networks have improved image generation by introducing a generator and a discriminator [1]. The generator tries to create realistic images, while the discriminator tries to detect fake images. Through this process, the model learns the data distribution.

However, early models still require paired data. CycleGAN was proposed to solve this problem and can work with unpaired datasets.

CycleGAN Architecture

CycleGAN is built on a dual mapping framework that consists of two generators and two discriminators, enabling bidirectional translation between two image domains. One generator G learns a mapping from domain X to domain Y, while the other generator F performs the inverse mapping from domain Y back to domain X. This bidirectional design allows the model to learn cross-domain relationships without requiring paired training data, which significantly increases its applicability in real world scenarios.

Each domain is associated with a corresponding discriminator. The discriminator *DX* distinguishes between real images from domain *X* and generated images produced by *F*, while *DY* performs the same task for domain *Y*. The discriminators are trained to classify real and fake samples, whereas the generators are optimized to produce images that are indistinguishable from real ones. This adversarial learning process enables the model to approximate the underlying data distributions of both domains.

The generators are typically implemented using convolutional neural networks with residual learning mechanisms. A standard generator consists of an encoder, several residual blocks, and a decoder. The encoder extracts high-level feature representations from the input image, the residual blocks transform these representations while preserving important structural information, and the decoder reconstructs the translated image. This design helps maintain the global structure of the input while performing domain transformation.

Residual blocks play a critical role in improving model performance. By learning residual mappings instead of direct transformations, they simplify the optimization process and help preserve essential image content. This is particularly important in image translation tasks where structural consistency must be maintained.

Instance normalization is commonly used in the generator networks. It normalizes feature statistics for each image independently, which improves training stability and enhances the model's ability to generalize across different data distributions.

The discriminators are usually implemented as Patch Generative Adversarial Network (PatchGAN) classifiers. Instead of evaluating the entire image, PatchGAN operates on local patches, allowing it to focus on high-frequency details such as textures and edges. As a result, the generated images exhibit improved local realism and finer visual details.

A key component of CycleGAN is the cycle consistency mechanism. When an image is translated from domain *X* to domain *Y* using generator *G* and then mapped back to domain *X* using generator *F*, the reconstructed image should be close to the original input. This constraint enforces consistency between forward and backward mappings, prevents arbitrary transformations, and ensures that the learned mappings preserve the semantic content of the images [2].

Training Strategy

The training process of CycleGAN is based on alternating optimization between generators and discriminators. In each iteration, the generators first produce translated images, and then the discriminators evaluate whether these images are real or generated. This process continues until both parts reach a balance.

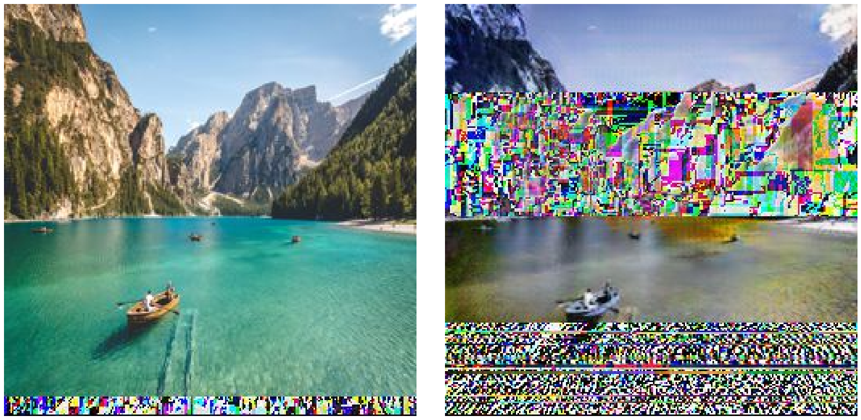
Another important technique is identity mapping. When an image from the target domain is input into the generator, the output should remain close to the input. This helps preserve color and brightness information.

In practice, the Adam optimizer is commonly used. Learning rate scheduling is also applied to improve convergence. Proper parameter tuning is important because this model can be unstable during training.

Experimental Results

To evaluate the performance of the CycleGAN model, experiments were conducted on several public datasets. The horse-to-zebra dataset is one of the most commonly used benchmarks. In this task, the model learns to convert horses into zebras while preserving the overall shape and structure of the original images.

Another dataset involves seasonal translation, where images are converted between summer and winter scenes. This task is more challenging because it requires significant changes in lighting conditions, color distribution, and texture details, while still maintaining the underlying scene structure.



a *b*

Figure 1. Example of seasonal image translation: *a* – summer;*b* – winter

As shown in Figure 1, the model effectively converts the summer image into a winter scene while preserving the overall structure of the original image.

Two evaluation metrics were used: structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). SSIM measures the similarity between images in terms of structure, while PSNR evaluates the overall reconstruction quality.

The experimental results show that CycleGAN is capable of generating visually realistic images while effectively preserving the main structural content of the input. Although its performance is slightly lower than that of supervised methods in some cases, it eliminates the need for paired training data, which represents a significant advantage in practical applications [3].

Applications and Limitations

CycleGAN has been widely used in various domains due to its ability to learn mappings from unpaired data. In medical imaging, it can be applied to translate between different imaging modalities such as MRI and CT, which improves data utilization and reduces the need for paired datasets. In computer graphics, CycleGAN enables style transfer tasks, such as converting photographs into artistic images. In autonomous driving, it can be used to enhance simulated data so that it better matches real-world environments, thereby improving model generalization.

The experimental results in this paper, including the horse-to-zebra and seasonal translation tasks, show that CycleGAN can generate visually realistic images while preserving the overall structure of the input. These findings are consistent with previous work, which demonstrates that the cycle consistency constraint helps preserve content during translation.

However, several limitations remain. First, when the difference between the two domains is large, the model may fail to preserve important semantic structures. This issue has been discussed in previous studies, where the mapping is learned without explicit semantic supervision [4].

Second, the model may produce visual artifacts, especially in regions with complex textures or fine details. This is a common limitation of adversarial training, where the objective focuses more on visual realism than structural accuracy.

In addition, CycleGAN provides limited control over the translation process. The mapping between domains is learned implicitly, making it difficult to control specific attributes or generate diverse outputs. This limitation has motivated later work on multimodal image translation [5] and contrastive learning-based approaches [6].

Overall, while CycleGAN provides a flexible solution for unpaired image-to-image translation and reduces the reliance on labeled data, its performance is still limited by the lack of semantic constraints and controllability. Addressing these issues remains an important direction for future research.

Conclusion

This paper presents an overview of image-to-image translation methods, with a focus on CycleGAN. Compared with supervised approaches such as Pix to Pix, CycleGAN can be trained using unpaired data, making it more flexible for real-world applications. The comparison results show that CycleGAN achieves competitive performance while reducing the need for labeled datasets.

However, as discussed in the previous section, CycleGAN still has several limitations. In particular, it may fail to preserve semantic consistency when the domain gap is large, and it may introduce visual artifacts in complex scenes. In addition, the lack of controllability limits its applicability in tasks that require precise transformations.

Future work can focus on improving semantic consistency and controllability. Possible directions include incorporating attention mechanisms, introducing additional constraints, or exploring more advanced models for image translation.

Overall, CycleGAN provides an effective solution for unpaired image-to-image translation, but further improvements are needed to enhance its robustness and applicability.

References

1. Goodfellow I., Pouget-Abadie J., Mirza M., et al. // Proc. of the Neural Information Processing Systems. 2014. P. 2672–2680.
2. Zhu J., et al. // Proc. of the IEEE International Conference on Computer Vision. 2017. P. 2223–2232.
3. Isola P., et al. // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 1125–1134.
4. Taigman Y., et al. // Proc. of the International Conference on Learning Representations. 2017.
5. Huang X., et al. // Proc. of the European Conference on Computer Vision. 2018. P. 172–189.
6. Park T., et al. // Proc. of the European Conference on Computer Vision. 2020. P. 319–345.